Signatures of selection acting on the innate immunity gene Toll-like receptor 2 (TLR2) during the evolutionary history of rodents

Tschirren, Barbara ; Raberg, Lars ; Westerdahl, Helena

Abstract: Patterns of selection acting on immune defence genes have recently been the focus of considerable interest. Yet, when it comes to vertebrates, studies have mainly focused on the acquired branch of the immune system. Consequently, the direction and strength of selection acting on genes of the vertebrate innate immune defence remain poorly understood. Here, we present a molecular analysis of selection on an important receptor of the innate immune system of vertebrates, the Toll-like receptor 2 (TLR2), across 17 rodent species. Although purifying selection was the prevalent evolutionary force acting on most parts of the rodent TLR2, we found that codons in close proximity to pathogen- binding and TLR2–TLR1 heterodimerization sites have been subject to positive selection. This indicates that parasite-mediated selection is not restricted to acquired immune system genes like the major histocompatibility complex, but also affects innate defence genes. To obtain a comprehensive understanding of evolutionary processes in host–parasite systems, both innate and acquired immunity thus need to be considered.

DOI: https://doi.org/10.1111/j.1420-9101.2011.02254.x

Signatures of selection acting on the innate immunity gene Toll-like receptor 2 (TLR2) during the evolutionary history of rodents

Barbara Tschirren*, Lars Råberg & Helena Westerdahl

Molecular Ecology and Evolution Lab, Department of Biology, Lund University, Sölvegatan 37, 223 62 Lund, Sweden

***Correspondence:** Barbara Tschirren, Institute of Evolutionary Biology and Environmental Studies, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland; Tel: +41 44 635 47 77, Fax: +41 44 635 68 18, Email: barbara.tschirren@ieu.uzh.ch

Running title: Positive selection on TLR2 in rodents

Abstract

Patterns of selection acting on immune defence genes have recently been the focus of considerable interest. Yet, when it comes to vertebrates, studies have mainly focused on the acquired branch of the immune system. Consequently, the direction and strength of selection acting on genes of the vertebrate innate immune defence remain poorly understood. Here, we present a molecular analysis of selection on an important receptor of the innate immune system of vertebrates, the Toll-like receptor 2 (TLR2), across 17 rodent species. Although purifying selection was the prevalent evolutionary force acting on most parts of the rodent TLR2, we found that codons in close proximity to pathogen-binding and TLR2-TLR1 heterodimerisation sites have been subject to positive selection in rodents. This indicates that parasite-mediated selection is not restricted to acquired immune system genes like the Major histocompatibility complex (MHC), but also affects innate defence genes. To obtain a comprehensive understanding of evolutionary processes in host-parasite systems, both innate and acquired immunity thus need to be considered.

Key words: arms race, coevolution, disease ecology, host-parasite interactions, innate immunity, parasite-mediated selection, rodents, Toll-like receptor 2

Introduction

The strength and direction of parasite-mediated selection acting on hosts can vary in time and space as a result of host-parasite coevolution and / or variation in the composition of the parasite assemblage. As a result, parasites are often assumed to be important drivers of genetic diversification within and among host species (Hedrick, 2002; Woolhouse *et al.*, 2002). In line with this, a number of studies have demonstrated directional selection (i.e. selective sweeps that favour advantageous mutations and thereby results in divergence between species (Hughes, 1999)) or balancing selection (which maintains genetic diversity within species) on immune defence genes in both plants and animals (Tanaka & Nei, 1989; Tiffin & Moeller, 2006; Obbard *et al.*, 2009). In the case of plants and invertebrates, general patterns of selection on immune defence genes are beginning to emerge (Tiffin & Moeller, 2006). For example, a recent, comprehensive analysis of immune defence genes across *Drosophila* species revealed that pathogen recognition genes, rather than immune signalling or effector genes, are the primary targets of parasite-mediated selection (Sackton *et al.*, 2007). However, when it comes to vertebrates, it is still largely unclear which types of host defence genes are subject to positive selection from parasites (Tiffin & Moeller, 2006).

The vertebrate immune system consists of two parts, the innate and acquired immune defence. The innate branch of the immune system is ancient and has homologous components in both invertebrates and plants (Beutler, 2004). It provides a fast but generally non-specific defence against parasites. The acquired immune system, on the other hand, is characterized by high specificity and memory, and occurs exclusively in vertebrates (Cooper & Alder, 2006). Some acquired immune defence genes have been shown to be subject to strong selective sweeps and / or

59 balancing selection in a wide range of taxa, with the Major histocompatibility complex
60 (MHC) genes being exceptionally well studied in this context (Apanius *et al.*, 1997;
61 Hughes & Yeager, 1998; Piertney & Oliver, 2006; Wegner, 2008).

62 In contrast, patterns of selection acting on vertebrate innate immune genes are less
63 clear. As reviewed above, studies of selection on defence genes in plants and
64 invertebrates have shown that innate immunity genes, in particular genes involved in
65 pathogen recognition, can be subject to positive selection (Jiggins & Kim, 2006; Tiffin
66 & Moeller, 2006; Sackton *et al.*, 2007; Obbard *et al.*, 2009). Yet, in vertebrates the
67 presence of a highly specific defence, mediated by the acquired immune system,
68 might affect the role, and consequently the patterns of selection acting on their
69 innate immune defence. Indeed, it is often argued that because innate defence
70 receptors recognize conserved pathogen structures, they should be subject to
71 purifying (i.e. negative) rather than positive selection (Medzhitov & Janeway, 1997;
72 Mukherjee *et al.*, 2009).

73 One of the best-studied types of innate defence receptors are Toll-like receptors
74 (TLRs) (Vasselon & Detmers, 2002; Takeda *et al.*, 2003). TLRs form an extended
75 gene family, which has evolved by gene duplication (Zhou *et al.*, 2007). The general
76 structure of TLRs is characterized by an extracellular domain, which consists of a
77 number of leucine-rich repeats (LRRs) that are involved in pathogen recognition (Jin
78 *et al.*, 2007), a transmembrane region, and a highly conserved intracellular
79 Toll/Interleukin-1 receptor (TIR) domain (Kobe & Kajava, 2001; West *et al.*, 2006).
80 After binding to a pathogen, TLRs form homo- or heterodimers and initiate an
81 intracellular signalling cascade, which leads to the activation of transcription factors,
82 the production of pro-inflammatory cytokines and, ultimately, inflammation in the
83 infected host tissue (Medzhitov, 2001; Akira & Takeda, 2004). Most mammals have

10 – 12 different TLRs, each recognizing different ligands (Roach *et al.*, 2005). TLR2, for example, recognizes lipoproteins from cell walls of bacteria, whereas TLR3 targets double-stranded RNA of viruses (Garantziotis *et al.*, 2008).

Because Toll-like receptors, like MHC molecules, are involved in pathogen recognition, it has been suggested that they might be involved in coevolutionary processes with pathogens (Hughes & Friedman, 2008). Indeed, recent studies on primates have found evidence for positive selection on TLRs (Ferrer-Admetlla *et al.*, 2008; Hughes & Piontkivska, 2008; Wlasiuk *et al.*, 2009; Enard *et al.*, 2010). Yet, in agreement with the hypothesis that there is strong selection against functional change in innate immune genes in vertebrates, several studies in humans and livestock found that TLRs have been subject to purifying selection (Yilmaz *et al.*, 2005; Barreiro *et al.*, 2009; Mukherjee *et al.*, 2009; Seabury *et al.*, 2010). Hence, analyses of patterns of selection on TLRs have yielded surprisingly mixed results.

To shed further light on the strength and direction of selection acting on the vertebrate innate immune defence in general, and TLRs in particular, and to identify putatively positively selected TLR sites, we studied signatures of selection in the Toll-like receptor 2 (TLR2) gene across 17 rodent species. Genetic polymorphisms in TLR2 are associated with resistance to a variety of disease causing agents, including *Staphylococcus aureus*, *Mycobacterium tuberculosis*, *Mycobacterium leprae*, *Treponema pallidum*, and *Borrelia burgdorferi* in humans and lab mice (Cook *et al.*, 2004; Schröder *et al.*, 2005; Texereau *et al.*, 2005; Garantziotis *et al.*, 2008; Schröder *et al.*, 2008), which makes TLR2 a particularly good candidate for the study of selection acting on innate immune defence in wild populations.

Material and Methods

Samples

Tissue samples from wild populations of 15 rodent species (Table 1) were obtained using live-traps (Ugglan Special No1, Grahnab, Gnosjö, Sweden) or from colleagues working with the species. In addition, we used TLR2 sequences of *Mus musculus* (Rodentia, Murinae, NM_011905.3), *Rattus norvegicus* (Rodentia, Murinae, NM_198769.2) and *Cavia porcellus* (Rodentia, Caviinae, ENSCPOG00000025374) publicly available from GenBank and Ensembl.

Primer design

Murinae primers were designed using the *Mus musculus* (NM_011905.3) TLR2 sequence and the program Primer3 (Rozen & Skaletsky, 2000). For the Arvicolinae primers, we first aligned the *Mus musculus* and *Rattus norvegicus* TLR2 sequences and identified highly conserved regions. Primers designed to anneal to this conserved part of the TLR2 sequence were then used to amplify the homologous region in the bank vole (*Myodes glareolus*). Based on this partial bank vole TLR2 sequence, we designed new bank vole specific primers and performed 5'- and 3'-RACE (Rapid Amplification of C-DNA Ends) with total RNA from bank vole spleen using the SMART RACE cDNA Amplification Kit (Clontech, Mountain View, CA, USA). This technique allowed us to obtain the full-length cDNA sequence of the bank vole TLR2. Specific Arvicolinae primers were then designed based on this bank vole sequence using the program Primer3 (Rozen & Skaletsky, 2000).

DNA amplification and sequencing

Total DNA was extracted from the biopsies following the protocol of Laird *et al.* (1991). For each rodent species, we amplified the entire coding region of TLR2 in two overlapping amplicons. PCR reactions were performed in a total volume of 25 µl including 25 ng of total genomic DNA, 0.125 mM of each dNTP, 2.0 mM $MgCl_2$, 1x PCR Buffer (Applied Biosystems, Foster City, CA, USA), 1 mM of each primer, and 2.5 U AmpliTaq DNA polymerase (Applied Biosystems, Foster City, CA, USA) on a GeneAmp PCR Systems 9700 thermocycler (Applied Biosystems, Foster City, CA, USA). For the first amplicon we used the primer pairs Mur1TLR2F 5'-MRSGTCAAATCTCAGAGGATG-3' and Mur1TLR2R 5'-GAGTYACACMKRTAGCTGTCTG-3' for the Murinae, and Arv1TLR2F 5'-CGTGTTCTGTGGACCTTGTG-3' and Arv1TLR2R 5'-CTAACATCCAGCACCTCCAG-3' for the Arvicolinae. For the second amplicon, we used the primer pairs Mur2TLR2F 5'-CAAACTGRAGACTYTGGAAGC-3' and Rod2TLR2R 5'-GAACCTAGGACTTTATTGCAGTTCTC-3' for the Murinae, and Arv2TLR2F 5'-CTTGACATCAGCCGGAACAG-3' and Rod2TLR2R for the Arvicolinae. Additional internal primers were used for sequencing.

The PCR protocol included an initial denaturation step at 94 °C for 5 minutes, followed by 35 cycles of denaturation at 94 °C for 30 seconds, annealing at 56 °C for 30 seconds and extension at 72 °C for 150 seconds. The program ended with a final extension step at 72 °C for 10 minutes. The PCR products were sequenced in both directions on an ABI Prism 3730 capillary sequencer (Applied Biosystems, Foster City, CA, USA) using Big Dye terminator v3.1 chemistry (Applied Biosystems, Foster City, CA, USA).

156 *Sequence analysis*

157 Sequences were processed, assembled and aligned using ClustalW in the program
158 Geneious 5.0. (Drummond *et al.*, 2009). Individual alignments were checked and
159 improved by eye. Consensus sequences were created and all inconsistencies and
160 polymorphisms were examined by eye. The intracellular region of TLR2 is very
161 highly conserved across taxa, we never observed a mismatch between overlapping
162 amplicons and there was no indication that we amplified more than one locus.
163 Together, this strongly suggests that sequences obtained from the different rodent
164 species indeed represent true homologues. The full alignments of the rodent TLR2
165 sequences are presented in Supplementary material 1. The length of the predicted
166 TLR2 protein was 782 – 784 amino acids in all sequenced rodent species.
167 Sequences were submitted to NCBI GenBank (see Table 1 for accession numbers).

168

169 *Tests of selection*

170 To test for signatures of past selection in the TLR2 DNA sequence of rodents, we
171 performed a number of standard tests (see e.g. Yang & Bielawski, 2000; Nielsen,
172 2005). The established structure and inferred function of the TLR2 protein (Xu *et al.*,
173 2000; Gautam *et al.*, 2006; Jin *et al.*, 2007) allowed us to make *a priori* predictions
174 about which sites are expected to evolve under positive selection (i.e. sites involved
175 in pathogen recognition), and thus to put our results into a functional perspective.

176 First, to identify regions that may have evolved under positive selection, we
177 performed a sliding window analysis of the ratio (K_A / K_S or ω) of non-synonymous
178 substitutions per non-synonymous site (K_A) to synonymous substitutions per

179 synonymous site (K_S) along the TLR2 gene of Murinae, Myodini and Arvicolini,
180 compared to the outgroup *Cavia porcellus* using DnaSP 5 (Librado & Rozas, 2009).
181 The Jukes-Cantor correction (JC) was applied when calculating K_A and K_S. The ratio
182 was calculated for the entire coding region of TLR2 and was averaged for a window
183 of 30 bp, with a step size of 10 bp. A ratio (ω) > 1 is indicative of positive selection
184 (i.e. promoting change at the amino acid level), whereas ω < 1 is indicative of
185 purifying selection (Nielsen, 2005).

186 Second, we used the phylogeny-based maximum likelihood analysis of ω as
187 implemented in the program CODEML of the package PAML 4.3 (Yang, 1997) to
188 statistically test for positive selection acting on TLR2 codons. We generated log
189 likelihood values for models where ω is allowed to vary among sites within the
190 interval 0 – 1 (neutral models), and for models, which allow ω to be > 1 for some
191 sites (selection models) following Yang *et al.* (2000). First, we tested if ω differs
192 among sites by comparing model M0, which assumes a constant ω across all sites to
193 model M3, which allows ω to vary among sites. To formally test for the presence of
194 sites evolving under positive selection, we then compared a nearly neutral model of
195 ω variation (M1a) to a model that allows for positive selection (M2a) (Wong *et al.*,
196 2004; Yang *et al.*, 2005), and a neutral model M7, which estimates ω with a beta
197 distribution over the interval 0 – 1 to a selection model M8, which additionally allows
198 for positively selected sites (ω > 1) (Yang *et al.*, 2000; Yang & Nielsen, 2002). A beta
199 distribution of ω has been suggested to more accurately reflect the distribution of ω
200 among sites in biological data (Friedman & Hughes, 2007). Furthermore, we
201 compared beta model M8a (a special case of M8, which fixes ω of the highest site
202 class to 1) to M8 (Swanson *et al.*, 2003). We compared the models using likelihood

ratio tests. Twice the log-likelihood difference ($2\Delta\ell$) between models follows a χ^2 distribution with the degrees of freedom equal to the difference in the number of parameters between the models (Yang & Nielsen, 2002). All models were run multiple times with different starting values for ω to ensure the correct estimation of the model parameters. The unrooted tree input file (Supplementary material 2) for these analyses was created by calculating a distance matrix in the program DNAdist and a tree file in the program NEIGHBOR implemented in the package PHYLIP (Felsenstein, 2005).

Finally, we used Empirical Bayes approaches implemented in CODEML (Yang, 1997) to infer which sites of the TLR2 sequence may have evolved under positive selection (Yang *et al.*, 2005). This algorithm estimates for each site the posterior probability of belonging to one of three site classes: low ω, intermediate ω and high ω. We considered two different approaches to determine sites under selection, the naive-empirical Bayes and the Bayes-empirical Bayes method (Yang *et al.*, 2005). Positive selection was interfered if the posterior probability of $\omega > 1$ for a site was ≥ 0.95.

Results

The sliding window analysis of the ratio K_A / K_S along the entire TLR2 coding region produced a strong peak in LRRs 10 – 11. This signal of positive selection was repeatable across Murinae, Myodini and Arvicolini (Fig. 1). A second prominent K_A / K_S peak was observed in LRR 14 in Murinae and Arvicolini, but not in Myodini (Fig. 1). A high K_A / K_S can be the result of high K_A, low K_S, or a combination of both. To investigate what caused the peaks we observed, we therefore calculated K_A and K_S

227 separately for each of the peak regions and compared these values to the average
228 K_A and K_S values over different functional regions of TLR2 (Table 2). It turned out
229 that both peaks were caused by a combination of high K_A and low K_S. To evaluate
230 how extreme the K_A and K_S at the observed peaks were, we generated 100 random
231 windows of 30 bp (i.e. the window size of the sliding window analysis) along TLR2,
232 and calculated K_S and K_A for each window (Supplementary material 3). In this
233 generated dataset we did neither observe K_A values that were as high, nor K_S values
234 that were as low as at the two observed peaks obtained in the sliding window
235 analysis (Table 2, Supplementary material 3), indicating that the probability of
236 observing such extreme values by chance is < 1%.

237 In addition to the sliding window analysis, phylogeny-based maximum
238 likelihood approaches (Yang, 1997) provided evidence that positive selection has
239 acted on codons of the rodent TLR2 sequence. We detected significant ω
240 heterogeneity along the TLR2 sequence (Model 0 vs. 3: $\chi^2_4 = 200.32$, $P < 0.001$) with
241 68% of the sites evolving under strong purifying selection ($\omega = 0.06$), 32% of the
242 sites evolving neutrally ($\omega = 0.88$), and 0.5% of the sites evolving under strong
243 positive selection ($\omega = 4.8$) (Table 3). A first selection test, in which we compared a
244 nearly neutral model (M1a, two site classes) with a selection model (M2a, three site
245 classes), did not provide statistical support for positive selection ($\chi^2_2 = 0$, $P = 1$, Table
246 3). However, when comparing models that estimate ω with a beta distribution the
247 selection model M8 (eleven site classes), which allows $\omega > 1$, performed significantly
248 better than the neutral model M7 (ten site classes), which restricts ω to the interval 0
249 – 1 ($\chi^2_2 = 6.66$, $P = 0.036$, Table 3). Furthermore, the comparison of model M8a vs.
250 M8 (both eleven site classes; ($\chi^2_1 = 4.34$, $P = 0.037$) provided evidence that ω of the

highest site class is significantly > 1, and thus that parts of TLR2 are evolving under positive selection. Selection models did not perform significantly better than neutral models when analysing Murinae, Arvicolini, and Myodini separately (results not shown).

To identify specific sites of TLR2 that have likely evolved under positive selection, we used Empirical Bayes approaches (Yang *et al.*, 2005). Even though both model M8 and model M3 indicated that 0.5 – 0.6% of all TLR2 sites evolve under positive selection ($\omega = 4.6$, Table 3), which corresponds to 3 – 5 codons, we had the statistical power to identify only one site, amino acid 354, as being positively selected with a posterior probability ≥ 0.95 (Table 3, Fig. 2).

Discussion

Unlike in plants and invertebrates, where it is well documented that positive selection has shaped innate immunity genes (Jiggins & Kim, 2006; Tiffin & Moeller, 2006; Sackton *et al.*, 2007; Obbard *et al.*, 2009), it is often assumed that in vertebrates purifying selection is the predominant selective force acting on the innate branch of the immune system (Yilmaz *et al.*, 2005; Barreiro *et al.*, 2009; Mukherjee *et al.*, 2009; Seabury *et al.*, 2010). Nonetheless, a few recent studies have demonstrated that also in vertebrates positive selection has acted on functionally relevant sites of innate defence genes (Ferrer-Admetlla *et al.*, 2008; Hughes & Piontkivska, 2008; Wlasiuk *et al.*, 2009; Enard *et al.*, 2010). In agreement with these latter studies, we found indication that positive selection has shaped parts of the rodent TLR2 gene.

273 Positive selection is predicted to act on regions of immune genes that are involved in
274 pathogen recognition (e.g. Hughes *et al.*, 1990; Hedrick, 2002; Hughes & Friedman,
275 2008). In agreement with this prediction, the sliding window analysis of K_A / K_S along
276 the TLR2 coding region produced a pronounced signal of positive selection in LRR
277 10 – 11. This signal of selection was repeatable across all three analysed taxonomic
278 groups. A second peak was observed in LRR 14 in Murinae and Arvicolini. The
279 LRRs 10 – 11 regions contain 10 to 13 putative pathogen-binding sites in humans
280 and lab mice (Jin *et al.*, 2007), and variation in these binding sites might directly
281 influence the pathogen-binding capacity or specificity of the receptor. LRR14, on the
282 other hand, contains sites that are involved in the heterodimerisation of TLR2 with
283 TLR1 (Jin *et al.*, 2007). Interestingly, the common human TLR2 variant Arg753Gln,
284 which is associated with resistance to late stage Lyme disease (Schröder *et al.*,
285 2005), also affects TLR2 heterodimerisation (Gautam *et al.*, 2006). This indicates
286 that even though heterodimerisation sites are not directly involved in pathogen-
287 binding, they might indirectly influence the pathogen-binding ability of the receptor by
288 changing the conformation of the TLR2 – TLR1 hererodimer structure (Gautam *et al.*,
289 2006).

290 Sliding window analyses of K_A / K_S have been criticised (see e.g. Hughes &
291 Friedman, 2008; Nozawa *et al.*, 2009; Wolf *et al.*, 2009) because K_A / K_S peaks could
292 reflect regions with particularly low K_S rather than high K_A, something which can
293 occur by chance given the high variance of K_S (Parmley & Hurst, 2007). In such
294 cases, K_A / K_S peaks will not be indicative of positive selection. A detailed
295 examination of our data, however, showed that stochastic variation of K_S alone
296 cannot explain the observed K_A / K_S peaks. Rather the K_A / K_S peaks are a result of
297 both exceptionally high K_A as well as exceptionally low K_S. An outlier analysis

298 demonstrated that such extreme K_A and K_S values would be expected by chance in <
299 1%. The reason for the low K_S is not entirely clear, but one potential explanation is
300 that selective sweeps (indicated by the high K_A), in combination with a high
301 recombination rate, have depleted these particular regions of neutral variation. In any
302 case, the fact that the K_A / K_S peaks are at least partly a result of high K_A, in
303 combination with the repeatability of the peaks across the analysed clades, and the
304 location of the peaks in the predicted regions of TLR2 (Jin *et al.*, 2007) provides
305 strong evidence that the observed K_A / K_S peaks are indicative of positive selection
306 rather than chance events.

307 The notion that positive selection shapes parts of the rodent TLR2 was further
308 corroborated by phylogeny-based maximum likelihood approaches. Although 68% of
309 the TLR2 sequence was estimated to evolve under strong purifying selection and
310 another 32% evolving neutrally, 0.5 – 0.6% of all sites (which corresponds to 3 – 5
311 amino acids) were found to be shaped by positive selection. Using empirical Bayes
312 approaches we were able to identify one of these positively selected sites, amino
313 acid 354. This site is located in LRR 12 in close proximity to both pathogen-binding
314 and heterodimerisation sites (Jin *et al.*, 2007). The nonsynonymous mutations
315 observed at this site have led to pronounced changes in the amino acids' physico-
316 chemical properties (i.e. polarity, charge, volume; Supplementary material 1), which
317 are thought to be important in the determination of protein structure (Grantham,
318 1974; Miyata *et al.*, 1979). Such radical changes are expected if natural selection
319 promotes functional change of proteins, but they are incompatible with the
320 substitution pattern predicted under relaxed purifying selection (see e.g. Nielsen,
321 2005). Remarkably, LRR12 showed a very low K_A / K_S ratio in the sliding window
322 analysis, indicating that codons in close proximity to amino acid 354 are subject to

strong purifying selection, and thus that patterns of selection can differ markedly on a very small scale.

To conclude, our results indicate that purifying selection is not the only evolutionary force that has shaped the rodent TLR2 sequence, as a small number of TLR2 codons have evolved under positive selection. There are numerous examples of single nucleotide polymorphisms having pronounced evolutionary consequences, for example in genes involved in pathogen virulence (Brault *et al.*, 2007) or pigmentation (Hoekstra *et al.*, 2004; Rosenblum *et al.*, 2004; Linnen *et al.*, 2009). Since the positively selected sites in the TLR2 sequence were in close proximity to pathogen-binding and TLR heterodimerisation sites (Jin *et al.*, 2007), it is thus plausible that functional changes in these regions may have direct consequences for host-pathogen interactions. So far, ecological and evolutionary research into the interactions between vertebrate hosts and their parasites has almost exclusively focused on the acquired branch of the immune system (i.e. associations between host MHC polymorphisms and parasite resistance (Paterson *et al.*, 1998; Wegner *et al.*, 2003; Meyer-Lucht & Sommer, 2005; Oliver *et al.*, 2009)), whereas the host's innate immune defence has received considerably less attention. Associating putatively positively selected amino acid differences in innate immunity genes with parasite resistance and tolerance in wild vertebrate populations might thus be a fruitful endeavour to obtain a more comprehensive understanding of evolutionary processes in natural host-parasite systems.

Furthermore, future studies will need to examine whether patterns of selection on TLR2 are exceptional, or if also other members of the Toll-like receptor gene family,

and other innate immunity genes in general, have been subject to positive selection during the evolutionary past, and if differences in selection patterns can be associated with functional difference between immune genes. New technologies (such as next generation sequencing) will greatly facilitate such comparative studies.

Acknowledgements

We are grateful to Hitoshi Suzuki, Zbyszek Boratynski, Gerald Heckel, Anna Lindholm, Melanie Monroe, Mohammad Nafi S. Al-Sabi, Alice Remy and Peter Wandeler for providing samples from their study populations, Kristin Scherman for providing bank vole RNA samples and Martin Stervander for help with MrBayes. We thank Staffan Bensch and Bengt Hansson for discussion, and Phil Hedrick, Erik Postma, Pedro Vale and two anonymous reviewers for comments on earlier versions of the manuscript. The project was funded by the Swedish Research Council (grants 621-2206-2876 and 621-2006-4551 to HW and LR). BT was supported by a Swiss National Science Foundation Postdoctoral Fellowship (PA0033-121466).

Table legends

Table 1 Origin of rodent samples, number of sequenced individuals (N) and NCBI GenBank accession numbers. *Sequences obtained from NCBI GenBank.

Table 2 Non-synonymous substitutions per non-synonymous site (K_A) and synonymous substitutions per synonymous site (K_S) averaged over the whole TLR2 coding region, the leucine-rich repeat 9 – 14 region (LRR 9 – 14), where most of the pathogen-binding and dimerisation sites are situated (Jin *et al.* 2007), the intracellular TIR domain and peak 1 and peak 2 identified in the sliding-window analysis (Fig. 1) for all species combined and the Murinae, Myodini and Arvicolini separately. Peak 1 and 2 are characterised by both high K_A and low K_S.

Table 3 Results of PAML site-models and positively selected sites identified by empirical Bayes approaches. Parameters are p_0 = proportion of sites where $\omega < 1$, p_1 = proportion of sites where $\omega = 1$ and p_2 = proportion of sites where $\omega > 1$ (selection models only). For models M7, M8 and M8a, p and q represent parameters of the beta distribution. Positive selection was inferred if the posterior probability of $\omega > 1$ for a site was 0.95 or higher (bold). Sites with a posterior probability of $\omega > 1$ between 0.50 – 0.949 are also shown (italic). We considered two different approaches to determine sites under selection, the naive-empirical Bayes and the Bayes-empirical Bayes method (Yang *et al.*, 2005). A site was included if one or both of the approaches gave statistical support for the site. Amino acids correspond to the *Mus musculus* sequence.

389 **Figure legends**

390

391 **Fig. 1** Signatures of positive selection on TLR2. Sliding window analysis of the ratio
392 of non-synonymous substitutions per non-synonymous site (K_A) to synonymous
393 substitutions per synonymous site (K_S) along the entire coding region of the TLR2
394 gene of A) all species combined, and B) Murinae, C) Myodini and D) Arvicolini
395 separately compared to *Cavia porcellus* (window size 30 bp, step size 10 bp). A ratio
396 > 1 is indicative of positive selection. Note the different scales on the y-axes. The
397 schema of the functional domains and the putative positions of the pathogen-binding
398 and TLR2 – TLR1 heterodimerisation sites are based on Gautam *et al.* (2006), Jin *et*
399 *al.* (2007), and Jann *et al.* (2008). The positively selected site (PSS) identified by
400 empirical Bayes approaches is shown in red. LRR: leucine-rich repeat, TM:
401 transmembrane domain, TIR: Toll/ Interleukin-1 receptor domain.

402

403 **Fig. 2** Expected ω of codons along the TLR2 sequence. Estimates are based on
404 Naive empirical Bayes probabilities for three classes.

405

406 **References**

407 Akira, S. & Takeda, K. 2004. Toll-like receptor signalling. *Nat. Rev. Immunol.* **4**: 499-
408 511.

409 Apanius, V., Penn, D., Slev, P.R., Ruff, L.R. & Potts, W.K. 1997. The nature of
410 selection on the Major Histocompatibility Complex. *Crit. Rev. Immunol.* **17**:
411 179-224.

412 Barreiro, L.B., Ben-Ali, M., Quach, H., Laval, G., Patin, E., Pickrell, J.K., Bouchier,
413 C., Tichit, M., Neyrolles, O., Gicquel, B., Kidd, J.R., Kidd, K.K., Alcais, A.,
414 Ragimbeau, J., Pellegrini, S., Abel, L., Casanova, J.L. & Quintana-Murci, L.
415 2009. Evolutionary dynamics of human Toll-like receptors and their different
416 contributions to host defense. *PLoS Genet.* **5**: e1000562.
417 doi:10.1371/journal.pgen.1000562.

418 Beutler, B. 2004. Innate immunity: an overview. *Mol. Immunol.* **40**: 845-859.

419 Brault, A.C., Huang, C.Y.H., Langevin, S.A., Kinney, R.M., Bowen, R.A., Ramey,
420 W.N., Panella, N.A., Holmes, E.C., Powers, A.M. & Miller, B.R. 2007. A single
421 positively selected West Nile viral mutation confers increased virogenesis in
422 American crows. *Nat. Genet.* **39**: 1162-1166.

423 Cook, D.N., Pisetsky, D.S. & Schwartz, D.A. 2004. Toll-like receptors in the
424 pathogenesis of human disease. *Nat. Immunol.* **5**: 975-979.

425 Cooper, M.D. & Alder, M.N. 2006. The evolution of adaptive immune systems. *Cell*
426 **124**: 815-822.

427 Drummond, A.J., Ashton, B., Cheung, M., Heled, J., Kearse, M., Moir, R., Stones-
428 Havas, S., Thierer, T. & Wilson, A. (2009) Geneious v4.6. pp. Available from
429 http://www.geneious.com.

Enard, D., Depaulis, F. & Crollius, H.R. 2010. Human and non-human primate genomes share hotspots of positive selection. *PLoS Genet.* **6**: e1000840. doi:10.1371/journal.pgen.1000840.

Felsenstein, J. (2005) PHYLIP (Phylogeny Inference Package) version 3.69. Distributed by the author. Department of Genome Sciences, University of Washington Seattle.

Ferrer-Admetlla, A., Bosch, E., Sikora, M., Marques-Bonet, T., Ramirez-Soriano, A., Muntasell, A., Navarro, A., Lazarus, R., Calafell, F., Bertranpetit, J. & Casals, F. 2008. Balancing selection is the main force shaping the evolution of innate immunity genes. *J. Immunol.* **181**: 1315-1322.

Friedman, R. & Hughes, A.L. 2007. Likelihood-ratio tests for positive selection of human and mouse duplicate genes reveal nonconservative and anomalous properties of widely used methods. *Mol. Phylogenet. Evol.* 42: 388-393.

Garantziotis, S., Hollingsworth, J.W., Zaas, A.K. & Schwartz, D.A. 2008. The effect of Toll-like receptors and Toll-like receptor genetics in human disease. *Ann. Rev. Med.* **59**: 343-359.

Gautam, J.K., Ashish, Comeau, L.D., Krueger, J.K. & Smith, M.F. 2006. Structural and functional evidence for the role of the TLR2 DD loop in TLR1/TLR2 heterodimerization and signaling. *J. Biol. Chem.* **281**: 30132-30142.

Grantham, R. 1974. Amino-acid difference formula to help explain protein evolution. *Science* **185**: 862-864.

Hedrick, P.W. 2002. Pathogen resistance and genetic variation at MHC loci. *Evolution* **56**: 1902-1908.

453 Hoekstra, H.E., Drumm, K.E. & Nachman, M.W. 2004. Ecological genetics of
454 adaptive color polymorphism in pocket mice: geographic variation in selected
455 and neutral genes. *Evolution* **58**: 1329-1341.
456 Hughes, A.L., Ota, T. & Nei, M. 1990. Positive Darwinian selection promotes charge
457 profile diversity in the antigen-binding cleft of class I Major Histocompatibility
458 Complex molecules. *Mol. Biol. Evol.* **7**: 515-524.
459 Hughes, A.L. & Yeager, M. 1998. Natural selection at Major Histocompatibility
460 Complex loci of vertebrates. *Annu. Rev. Genet.* **32**: 415-435.
461 Hughes, A.L. 1999. *Adaptive Evolution of Genes and Genomes*. Oxford University
462 Press, New York.
463 Hughes, A.L. & Friedman, R. 2008. Codon-based tests of positive selection, branch
464 lengths, and the evolution of mammalian immune system genes.
465 *Immunogenetics* **60**: 495-506.
466 Hughes, A.L. & Piontkivska, H. 2008. Functional diversification of the Toll-like
467 receptor gene family. *Immunogenetics* **60**: 249-256.
468 Jann, O.C., Werling, D., Chang, J.S., Haig, D. & Glass, E.J. 2008. Molecular
469 evolution of bovine Toll-like receptor 2 suggests substitutions of functional
470 relevance. *BMC Evol. Biol.* **8**: doi:10.1186/1471-2148-8-288.
471 Jiggins, F.M. & Kim, K.W. 2006. Contrasting evolutionary patterns in Drosophila
472 immune receptors. *J. Mol. Evol.* **63**: 769-780.
473 Jin, M.S., Kim, S.E., Heo, J.Y., Lee, M.E., Kim, H.M., Paik, S.G., Lee, H.Y. & Lee,
474 J.O. 2007. Crystal structure of the TLR1-TLR2 heterodimer induced by
475 binding of a tri-acylated lipopeptide. *Cell* **130**: 1071-1082.
476 Kobe, B. & Kajava, A.V. 2001. The leucine-rich repeat as a protein recognition motif.
477 *Curr. Opin. Struct. Biol.* **11**: 725-732.

478 Laird, P.W., Zijderveld, A., Linders, K., Rudnicki, M.A., Jaenisch, R. & Berns, A.
479 1991. Simplified mammalian DNA isolation procedure. *Nucleid Acids Res.* **19**:
480 4293.
481 Librado, P. & Rozas, J. 2009. DnaSP v5: A software for comprehensive analysis of
482 DNA polymorphism data. *Bioinformatics* **25**: 1451-1452.
483 Linnen, C.R., Kingsley, E.P., Jensen, J.D. & Hoekstra, H.E. 2009. On the origin and
484 spread of an adaptive allele in deer mice. *Science* **325**: 1095-1098.
485 Medzhitov, R. & Janeway, C.A. 1997. Innate immunity: The virtues of a nonclonal
486 system of recognition. *Cell* **91**: 295-298.
487 Medzhitov, R. 2001. Toll-like receptors and innate immunity. *Nat. Rev. Immunol.* **1**:
488 135-145.
489 Meyer-Lucht, Y. & Sommer, S. 2005. MHC diversity and the association to nematode
490 parasitism in the yellow-necked mouse (*Apodemus flavicollis*). *Mol. Ecol.* **14**:
491 2233-2243.
492 Miyata, T., Miyazawa, S. & Yasunaga, T. 1979. Two types of amino acid
493 substitutions in protein evolution *J. Mol. Evol.* **12**: 219-236.
494 Mukherjee, S., Sarkar-Roy, N., Wagener, D.K. & Majumder, P.P. 2009. Signatures of
495 natural selection are not uniform across genes of innate immune system, but
496 purifying selection is the dominant signature. *Proc. Natl. Acad. Sci. U S A* **106**:
497 7073-7078.
498 Nielsen, R. 2005. Molecular signatures of natural selection. *Annu. Rev. Genet.* **39**:
499 197-218.
500 Nozawa, M., Suzuki, Y. & Nei, M. 2009. Reliabilities of identifying positive selection
501 by the branch-site and the site-prediction methods. *Proc. Natl. Acad. Sci. U S*
502 *A* **106**: 6700-6705.

503 Obbard, D.J., Welch, J.J., Kim, K.W. & Jiggins, F.M. 2009. Quantifying adaptive
504 evolution in the Drosophila immune system. *PLoS Genet.* **5**: e1000698.
505 doi:10.1371/journal.pgen.1000698.
506 Oliver, M.K., Telfer, S. & Piertney, S.B. 2009. Major histocompatibility complex
507 (MHC) heterozygote superiority to natural multi-parasite infections in the water
508 vole (*Arvicola terrestris*). *Proc. R. Soc. Lond. B* **276**: 1119-1128.
509 Parmley, J.L. & Hurst, L.D. 2007. How common are intragene windows with K-A > K-
510 S owing to purifying selection on synonymous mutations? *J. Mol. Evol.* **64**:
511 646-655.
512 Paterson, S., Wilson, K. & Pemberton, J.M. 1998. Major histocompatibility complex
513 variation associated with juvenile survival and parasite resistance in a large
514 unmanaged ungulate population (*Ovis aries* L.). *Proc. Natl. Acad. Sci. U S A*
515 **95**: 3714-3719.
516 Piertney, S.B. & Oliver, M.K. 2006. The evolutionary ecology of the major
517 histocompatibility complex. *Heredity* **96**: 7-21.
518 Roach, J.C., Glusman, G., Rowen, L., Kaur, A., Purcell, M.K., Smith, K.D., Hood,
519 L.E. & Aderem, A. 2005. The evolution of vertebrate Toll-like receptors. *Proc.*
520 *Natl. Acad. Sci. U S A* **102**: 9577-9582.
521 Rosenblum, E.B., Hoekstra, H.E. & Nachman, M.W. 2004. Adaptive reptile color
522 variation and the evolution of the Mc1r gene. *Evolution* **58**: 1794-1808.
523 Rozen, S. & Skaletsky, H.J. 2000. Primer3 on the WWW for general users and for
524 biologist programmers. In: *Bioinformatics Methods and Protocols: Methods in*
525 *Molecular Biology*, (S., K. & S., M., eds.). pp. 365-386. Humana Press,
526 Totowa, NJ.

527 Sackton, T.B., Lazzaro, B.P., Schlenke, T.A., Evans, J.D., Hultmark, D. & Clark, A.G.
528 2007. Dynamic evolution of the innate immune system in Drosophila. *Nat.*
529 *Genet.* **39**: 1461-1468.
530 Schröder, N.W.J., Diterich, I., Zinke, A., Eckert, J., Draing, C., von Baehr, V.,
531 Hassler, D., Priem, S., Hahn, K., Michelsen, K.S., Hartung, T., Burmester,
532 G.R., Gobel, U.B., Hermann, C. & Schumann, R.R. 2005. Heterozygous
533 Arg753Gln polymorphism of human TLR-2 impairs immune activation by
534 *Borrelia burgdorferi* and protects from late stage Lyme disease. *J. Immunol.*
535 **175**: 2534-2540.
536 Schröder, N.W.J., Eckert, J., Stubs, G. & Schumann, R.R. 2008. Immune responses
537 induced by spirochetal outer membrane lipoproteins and glycolipids.
538 *Immunobiology* **213**: 329-340.
539 Seabury, C.M., Seabury, P.M., Decker, J.E., Schnabel, R.D., Taylor, J.F. & Womack,
540 J.E. 2010. Diversity and evolution of 11 innate immune genes in *Bos taurus*
541 *taurus* and *Bos taurus indicus* cattle. *Proc. Natl. Acad. Sci. U S A* **107**: 151-
542 156.
543 Swanson, W.J., Nielsen, R. & Yang, Q.F. 2003. Pervasive adaptive evolution in
544 mammalian fertilization proteins. *Mol. Biol. Evol.* **20**: 18-20.
545 Takeda, K., Kaisho, T. & Akira, S. 2003. Toll-like receptors. *Annu. Rev. Immunol.* **21**:
546 335-376.
547 Tanaka, T. & Nei, M. 1989. Positive Darwinian selection observed at the variable-
548 region genes of immunoglobulins *Mol. Biol. Evol.* **6**: 447-459.
549 Texereau, J., Chiche, J.D., Taylor, W., Choukroun, G., Comba, B. & Mira, J.P. 2005.
550 The importance of Toll-like receptor 2 polymorphisms in severe infections.
551 *Clin. Infect. Dis.* **41**: S408-S415.

552 Tiffin, P. & Moeller, D.A. 2006. Molecular evolution of plant immune system genes.
553 *Trends Genet.* **22**: 662-670.
554 Vasselon, T. & Detmers, P.A. 2002. Toll receptors: a central element in innate
555 immune responses. *Infect. Immun.* **70**: 1033-1041.
556 Wegner, K.M., Reusch, T.B.H. & Kalbe, M. 2003. Multiple parasites are driving Major
557 histocompatibility complex polymorphism in the wild. *J. Evol. Biol.* **16**: 224-
558 232.
559 Wegner, K.M. 2008. Historical and contemporary selection of teleost MHC genes:
560 did we leave the past behind? *J. Fish Biol.* **73**: 2110-2132.
561 West, A.P., Koblansky, A.A. & Ghosh, S. 2006. Recognition and signaling by Toll-like
562 receptors. *Annu. Rev. Cell Dev. Biol.* **22**: 409-437.
563 Wlasiuk, G., Khan, S., Switzer, W.M. & Nachman, M.W. 2009. A history of recurrent
564 positive selection at the Toll-like receptor 5 in primates. *Mol. Biol. Evol.* **26**:
565 937-949.
566 Wolf, J.B.W., Kunstner, A., Nam, K., Jakobsson, M. & Ellegren, H. 2009. Nonlinear
567 dynamics of nonsynonymous (d(N)) and synonymous (d(S)) substitution rates
568 affects inference of selection. *Gen. Biol. Evol.* **1**: 308-319.
569 Wong, W.S.W., Yang, Z.H., Goldman, N. & Nielsen, R. 2004. Accuracy and power of
570 statistical methods for detecting adaptive evolution in protein coding
571 sequences and for identifying positively selected sites. *Genetics* **168**: 1041-
572 1051.
573 Woolhouse, M.E.J., Webster, J.P., Domingo, E., Charlesworth, B. & Levin, B.R.
574 2002. Biological and biomedical implications of the co-evolution of pathogens
575 and their hosts. *Nat. Genet.* **32**: 569-577.

576 Xu, Y.W., Tao, X., Shen, B.H., Horng, T., Medzhitov, R., Manley, J.L. & Tong, L.
577 2000. Structural basis for signal transduction by the Toll/interleukin-1 receptor
578 domains. *Nature* **408**: 111-115.
579 Yang, Z. 1997. PAML: a program package for phylogenetic analysis by maximum
580 likelihood. *CABIOS* **15**: 555-556.
581 Yang, Z.H. & Bielawski, J.P. 2000. Statistical methods for detecting molecular
582 adaptation. *Trends Ecol. Evol.* **15**: 496-503.
583 Yang, Z.H., Nielsen, R., Goldman, N. & Pedersen, A.M.K. 2000. Codon-substitution
584 models for heterogeneous selection pressure at amino acid sites. *Genetics*
585 **155**: 431-449.
586 Yang, Z.H. & Nielsen, R. 2002. Codon-substitution models for detecting molecular
587 adaptation at individual sites along specific lineages. *Mol. Biol. Evol.* **19**: 908-
588 917.
589 Yang, Z.H., Wong, W.S.W. & Nielsen, R. 2005. Bayes empirical Bayes inference of
590 amino acid sites under positive selection. *Mol. Biol. Evol.* **22**: 1107-1118.
591 Yilmaz, A., Shen, S.X., Adelson, D.L., Xavier, S. & Zhu, J.J. 2005. Identification and
592 sequence analysis of chicken Toll-like receptors. *Immunogenetics* **56**: 743-
593 753.
594 Zhou, H.J., Gu, J.Y., Lamont, S.J. & Gu, X. 2007. Evolutionary analysis for functional
595 divergence of the Toll-like receptor gene family and altered functional
596 constraints. *J. Mol. Evol.* **65**: 119-123.
597
598